Probabilistic Phase Unwrapping for Single-Frequency Time-of-Flight Range Cameras

Under Active Illumination Model

Ryan Crabb
Dept. of Computer Engineering
University of California, Santa Cruz
rcrabb@soe.ucsc.edu

Roberto Manduchi
Dept. of Computer Engineering
University of California, Santa Cruz
manduchi@soe.ucsc.edu

***Abstract*—This paper proposes a solution to the 2-D phase unwrapping problem, inherent to time-of-flight range sensing technology due to the cyclic nature of phase. Our method uses a single frequency capture period to improve frame rate and decrease the presence of motion artifacts encountered in multiple frequency solutions. We present a probabilistic framework that considers intensity image in addition to the phase image. The phase unwrapping problem is cast in terms of global optimization of a carefully chosen objective function. Comparative experimental results confirm the effectiveness of the proposed approach.**

Keywords-Phase Unwrapping; Time-of-Flight; Markov Random Field; Range Camera

I. INTRODUCTION

Time-of-flight depth sensing technology (ToF) has matured in the last decade into one of the more common methods of range measurement, among the ranks of LIDAR, stereo depth, and structured light. These range sensing devices have become commercially available by manufacturers such as Mesa Imaging [**1**], Canesta [**2**], 3DV [**3**], and most notably Microsoft [**4**] (which purchased Canesta in …) with its second generation Kinect sensor for Xbox One. Because it is relatively inexpensive to produce and provides high capture rates, ToF range imaging is well suited for several applications despite a comparatively low accuracy and resolution.

Compared to other popular methods of range measurement, ToF offers its own advantages and disadvantages. Although it may be considered a subclass of scannerless lidar, ToF is distinct from typical scanning laser systems in that it has no moving parts and captures an entire 2D field simultaneously rather than scanning through the scene point by point. By capturing the entire scene at once, the ToF system is able to record depth at high frame rates, though because of the need to illuminate the entire field of view it does so at a sacrifice of maximum range, depth precision, spatial resolution, and power consumption. This makes ToF more suited towards close-range active applications (such as gesture-based user interface or interactive gaming) rather than precise 3D measurements of large static scenes. In comparison to stereo-based range imaging it requires much less computation, is not dependent upon a heavily textured surface, and not effected by ambient light. However, it is again limited in range and resolution as well as requiring more power for the active illumination. On the other hand, in comparison to structured light methods, ToF offers other advantages: because the setup is confocal it suffers no occlusion, and its resolution of measurement is independent of the hardware's geometry.

The term "time-of-flight" is in reference to the time it takes light to travel from the illumination source to the scene surface and back to the sensor. The illumination source is typically a near-infrared wavelength LED array or laser, placed very close (ideally, confocal) to the sensor. The returned light is imaged by a 2D CMOS array, making the hardware relatively inexpensive. Rather than measure the flight time of the light directly, the signal is modulated at a frequency f_m, and the phase shift θ of the recovered signal is an indication of the distance traveled. Specifically, the phase shift θ is related to the distance traveled by this equation:

$$2D = \frac{c\theta}{2\pi f_m} \qquad (1)$$

where D is the distance from the camera to the surface and c is the speed of light.

There are multiple ways in which the phase of the signal can be measured; the hardware uses in our measurements uses the method described in [**2**]. We assume that the intensity (*brightness*) image is available, in addition to phase measurement. This is usually the case in most ToF systems (e.g., the Canesta sensor).

Due to its cyclic nature, phase will wrap when it exceeds 2π, causing phase shift measurements to be ambiguous. For example, a measurement of 0.5π might signify an actual phase shift of 0.5π, 2.5π, 4.5π, etc. Formally, the phase θ in (1) can be decomposed as follows:

$$\theta = \Phi + 2\pi K \qquad (2)$$

where Φ is the observable wrapped phase and integer K represents the number of phase wraps. If the target surface is within the *maximally unambiguous range* of $c/2f_m$, then phase (and thus depth) recovery is straightforward. However, for scenes exceeding this depth, it is necessary to determine the phase wrap number K at each pixel in order to recover the correct depth of the scene. The task of determining the value of K for the collection of pixels in the (wrapped) phase image is known as the *2D phase unwrapping problem*, for which we present a new solution here.

A solution to the 2D phase unwrapping problem is to image the scene with two (or more) different modulation frequencies f_m [**2**]. However, this approach requires the scene to be static between the two exposures, and thus is ill suited to situations in which the scene is changing rapidly, or the camera is moving. The method presented here uses just one image, taken at a certain frequency f_m. It relies on both the wrapped phase measurements and the brightness image to estimate the true phase, and this the depth, of surfaces in the scene. The key innovation in this paper is in the derivation of the probability density function of the measured brightness from a surface illuminated by the sensor's light source, conditioned on the distance of the surface. The resulting expression is integrated in a Markov Random Field model defined on the wrapping numbers. Maximization of a suitable global objective function that includes a smoothness term defined on the estimated depth is performed via loopy belief propagation. Experiments with depth computation from phase and intensity images obtained from a single frequency ToF camera prototype show the effectiveness of our method when compared against the state of the art algorithm for single-frequency 2D phase unwrapping [**5**].

II. RELATED WORK

The phase unwrapping problem is not unique to time-of-flight depth imaging. The problem is encountered in technologies such as synthetic aperture radar (SAR) and magnetic resonance imaging (MRI), acoustic imaging, and microwave interferometry [**6**]. The phase unwrapping task may be in one, two or three dimensions (magnetic resonance imaging, as an example, images a volume and thus is a 3-dimensional phase unwrapping problem). Depth sensing ToF images occurs on a 2-D CMOS sensor and requires two-dimensional phase unwrapping.

Standard solutions to phase unwrapping [**6**] recover the unwrapped phase via path integration: the phase jump between adjacent pixels is determined, and the surface is recovered by first choosing an initial unwrapped phase value for the first pixel, and then spreading out over the rest of the image by adding the appropriate phase jump for each adjacent pixel. Using this approach, however, the phase is recoverable only up to a constant, which needs to be decided for the initially chosen pixel. Moreover, the solution may be path dependent, meaning that if the phase jumps between pixels are chosen independently, there is no guarantee that they will all be consistent (different paths leading to the same pixel may imply different changes in phase). This leads to an artifact called a *residue*: the case where the sum of phase jumps around a square of four neighboring pixels does not equal zero. Two families of algorithms for path integration are the *path following* methods and *minimum norm* methods. Path-following methods tend to either explicitly label discontinuities and choose paths that avoid them, or to grow the path (specifically, a tree) one step at a time without introducing residues. Minimum norm methods assign an L_p-norm penalty to phase jumps, and minimizes the sum of penalties over the entire grid. This and other methods are described in detail in [**6**]. While minimum norm methods can be very efficient to calculate, they do not guarantee that residues will be completely removed. Frey et al. address the residue issue by including a "zero-curl constraint" node in a Markov random field in [**7**], and this method was applied to time-of-flight imagery by Droeschel et al. [**8**]. In the solution proposed in this paper, the wrap number at each pixel is estimated directly, avoiding the possibility of residues entirely.

In theory, the need for phase unwrapping could be removed simply by increasing the unambiguous range. One method could be to decrease the modulation frequency; unfortunately, the measurement uncertainty increases proportionally with the unambiguous range [**2**]. Another way to increase the unambiguous range is through the use of multiple frequencies for the amplitude modulation of the active signal. This technique is adapted from INSAR technology [**9**] [**10**] and is a popular solution for ToF [**11**] [**12**] [**13**] [**14**]. If two frequencies are used, in successive or intermixed frame captures, then this pair can act as a single lower frequency. This will increase the unambiguous range by acting as an *effective frequency,* which is equivalent to the difference between the frequencies: $f_{eff} = \gcd(f_1, f_2)$. This is alternatively expressed as the effective unambiguous range and is the least common multiple of the unambiguous ranges of the two frequencies. One disadvantage of this approach is that it requires more power per frame to determine the distance. The key drawback to this approach, though, is that it requires more time to perform the multiple captures, rendering it more vulnerable to motion artifacts. This is especially troubling for tasks such as gesture detection, where the surface of interest is likely to be in motion. Some more recent work from the University of Waikato [**15**] uses multiple frequencies simultaneously, removing the need for some redundant samples. The overall capture time is reduced from consecutive multi-frequency, but still requires more time than a single capture.

Other solutions use just a single frequency, while incorporating additional information beyond wrapped

phase. One example is given by technology that combines a ToF camera with a traditional stereo camera pair. Gudmundsson et al. [**16**] use the ToF data to bootstrap the stereo pair. Beder et al. [**17**] combine the data to optimize local "patchlets" of the surface. Several others [**18**] [**19**] [**20**] fuse the data under a probabilistic framework using a Markov random field. Choi and Lee [**21**] exploit stereo principles using a pair of ToF cameras.

Some other recent work takes advantage of the intensity image measured by ToF cameras, along with the phase data. Our work fits squarely within this line of research. Böhme et al. use a shading constraint (à la shape from shading) in order to smooth noisy ToF depth images [**22**], however they assume the phase is already correctly unwrapped. In [**23**] the scene is first segmented by the depth edges, then the average intensity of each segment is analyzed to determine whether it falls into the unambiguous range, which is based on a manually set threshold. Choi et al. [**5**] first apply an EM optimization on the corrected intensity image [**24**] to classify each pixel as being within or outside the unambiguous range, which feeds the data term for a Markov random field optimization.

III. METHOD

We approach the phase unwrapping problem in a probabilistic framework, representing the number of phase wraps at each pixel as a discrete random variable. The novelty of our method is in the way we include the surface smoothness constraint as a prior in our model, and in the way we exploit the available brightness information. Rather than looking at the residue in a closed path [**7**], we model the difference in depth between two adjacent pixels as a normal random variable with zero mean and fixed variance. We also exploit the partial information about depth provided by the observed brightness. Brightness is a function of surface depth, slant, and albedo. Intuitively, bright pixels must correspond to nearby surfaces, whereas dark pixels may result from far away surfaces, or from surfaces that are close to the camera but have low albedo and/or large slant angle. This phenomenon was already observed by Choi et al., [**5**]. However, in contrast to the early-commitment segmentation approach of [**5**], we define a probabilistic model for the relation between brightness and distance, and use it as "data term" in a global optimization approach that includes the depth smoothness constraint mentioned above

1.1. Definitions and Problem Statement

For each frame we observe $O = \{\Phi, B\}$ consisting of a wrapped phase measurements [1] Φ and intensity (brightness) measurements B. We aim to recover the unwrapped phase θ at each pixel, given the observed wrapped phase Φ, by estimating the phase wrap number K (then $\theta = \Phi + 2\pi K$, see Eq. (2)). We model K as a uniform discrete random variable taking values between 0 and M. The maximum allowed number of wraps (M) is determined by the maximum distance of surfaces expected in the scene, or by the maximum distance at which a surface is visible (which depends on the power of the light source.) The phase wrap number assignment ("labeling") is performed using a maximum-a-posteriori (MAP) criterion[2]:

$$\begin{aligned} \widehat{K} &= \arg\max_K P(K|B,\Phi) \\ &= \arg\max_K p(B|K,\Phi)P(K|\Phi) \end{aligned} \qquad (3)$$

The first term in (3), $p(B|K,\Phi)$, represents the conditional likelihood of the brightness B given the wrapped phase Φ. A model for this quantity is derived in the following section. The second term is the posterior distribution on the label configuration K given the wrapped phase observation Φ. This term is derived from a smoothness prior on the depth field, as described in Sec. 1.3.

1.2. Illumination Model

The light intensity measured at each pixel can be attributed to three sources: light from the illumination source cast directly onto and reflected by the observed surface (direct reflection); light from the illumination source reflected indirectly off of neighboring surfaces (multipath reflection [**25**]); and ambient light. We consider only the direct reflection component of the measured brightness, which is usually the dominating component.

Our model assumes that illumination is from a point source, co-located with the camera. This means that the line of sight through a pixel coincides with the line of propagation of light illuminating the surface element imaged by that pixel. The illumination power, though, is typically not homogeneous, meaning that the transmitted power depends on the direction of propagation. The power distribution across pixels can be calibrated for a specific camera hardware, for example by measuring the brightness reflected off a planar surface of constant albedo, and compensating for the distance and angle of incidence of the surface element imaged by each pixel. After this calibration procedure, the light power along a propagation line through a pixel q can be represented by the brightness I_q measured at q from reflection against a surface element of albedo 1, at distance of 1 meter from the camera, orthogonal to the line of sight. If we assume

[1] We use capital letters to indicate the collection of values (field) over the whole image. To denote a specific value at a pixel q, we use a subscript notation, e.g. B_q.

[2] We will use the symbol P(·) to indicate probability distributions, and p(·) to indicate probability density functions.

that the visible surfaces are Lambertian, we can thus model the observed intensity B_q at a pixel as

$$B_q = \frac{I_q \cdot \rho \cdot \cos\beta}{D^2} \tag{4}$$

where ρ is the surface albedo, β is the angle of the surface normal with respect to the line of sight (slant), and D is the distance from the camera to the surface element imaged by p.

Hence, the brightness measured at a pixel depends on the (unknown) distance, slant, and albedo of the surface element. (Note that the measured brightness will be affected by Poisson noise; however, for the sake of simplicity, we will assume that it is noise-free.) We will make the following simplifying hypotheses: (a) the orientation of surfaces in the scene is modeled as a uniformly distributed random variable, which can be shown to result in a probability density function for the slant angle β equal to:

$$p(\beta) = 2\sin\beta\cos\beta \tag{5}$$

(b) the surface albedo ρ is a random variable uniformly distributed between 0 and 1; (d) the surface normal and albedo at a pixel are independent of the normal and albedo at other pixels. We thus obtain:

$$p(B|D) = \prod_q p(B_q|D_q) \tag{6}$$

with

$$p(B_q|D_q) = \tag{7}$$

$$= \int_0^1 \int_0^{\frac{\pi}{2}} p(B_q|D_q,\rho_q,\beta_q)\, p(\rho_q|\beta_q) p(\beta_q) d\beta_q d\rho_q$$

$$= \int_0^1 \int_0^{\frac{\pi}{2}} \delta\left(B_q - \frac{I_q \cdot \rho_q \cdot \cos\beta_q}{D_q^2}\right) \sin\beta_q \cos\beta_q \, d\beta_q d\rho_q$$

$$= \begin{cases} \frac{D_q^2}{I_q}\left[1 - \frac{B_q \cdot D_q^2}{I_q}\right], & \text{if } 0 \le B_q \le I_q/D_q^2 \\ 0, & \text{otherwise} \end{cases}$$

It is instructive to look at the posterior probability density function $p(D_q|B_q)$, which is easily obtained from $p(B_q|D_q)$ under the assumption of uniformly distributed D_q. As seen in Figure 1 and Figure 2, the spread of this density tightens with increasing values of B_p. This can be explained by the fact that a small brightness value at a pixel can be due to a surface being far away, but also to a surface having low albedo or being at a large slant angle. Hence, a dim pixel cannot provide much information about the surface distance alone, resulting in a large spread of $p(D_q|B_q)$. Conversely, *bright pixels can only be generated by a surface close to the camera*, with large albedo and small slant angle, resulting in a narrow spread

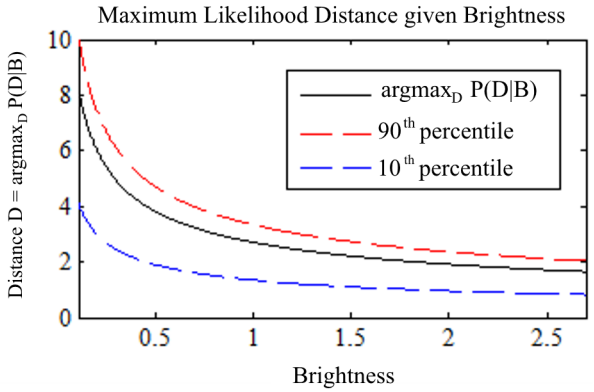

Figure 1. Outlining probability density of D for values of B. Black line shows most likely distance given measured brightness. Colored lines indicate the 10th and 90th percentiles. Curve is plotted over the observed range of Brightness and distance values, computed using a representative value for I.

of $p(D_q|B_q)$. This means that darker pixels have a high classification error rate, consistent with the increasing spread of the posterior density $p(D_q|B_q)$ with decreasing brightness. However, the misclassification rate is much smaller for brighter pixels; these pixels function as "anchor points" in the belief propagation step described in Sec. 1.4.

1.3. Smoothness Model

The "smoothness term" $P(K|\Phi)$ in (3) formalizes the notion that neighboring pixels are expected to have similar depths, and thus similar value of the unwrapped phase θ. We model the phase difference between neighboring pixels as a Gaussian with mean 0 and variance σ^2:

$$p(\theta_q - \theta_p) = N(\theta_q - \theta_p; 0, \sigma^2) \tag{8}$$

where $N(\cdot;\mu,\sigma^2)$ represents the normal pdf. Note that $\theta_q - \theta_p = 2\pi K_q + \Phi_q - 2\pi K_p - \Phi_p$. We will assume that: (a) the random variables Φ_q and Φ_p have uniform joint distribution; (b) the "dimension reduction"

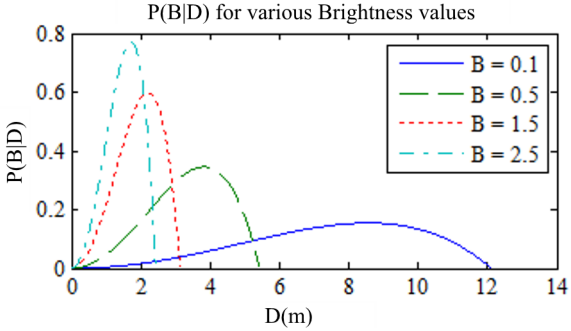

Figure 2. Conditional probability density for distance given measured brightness. For very bright values, surface must be nearby. For dim, it is more spread over the range.

$\Phi_q - \Phi_p$ is sufficient, meaning that $P(K_q - K_p|\Phi_q, \Phi_p) = P(K_q - K_p|\Phi_q - \Phi_p)$; (c) the joint statistical description of the variables $(\Phi_q - \Phi_p)$ and $(K_q - K_p)$ is fully represented by their sum: $p(\Phi_q - \Phi_p, K_q - K_p) = p(\theta_q - \theta_p)$. Under these simplifying conditions, it is easy to prove that

$$P(K_q - K_p|\Phi_q, \Phi_p) \propto N(K_q + \Phi_q - K_p - \Phi_p; 0, \sigma^2) \quad (9)$$

This expression can be used to establish a "factor potential" between neighboring pixels in a MRF labeling, as discussed in the next section.

The wrapping number could be inferred based on this smoothness term alone, using the global optimization framework described in the next section but neglecting the brightness term. Note that in this case the solution can be computed only up to a constant wrapping number (since adding any constant wrapping number does not change the smoothness term).

In closing this section, note that a robustified version of the smoothness term in (8) could be used to account for depth discontinuities. In our experiments, however, this did not lead to any noticeable advantage.

1.4. Global Optimization

We estimate the label configuration $\widehat{K}$ that maximizes $P(K|B, \Phi)$ in (3) by defining a global objective function $E(K)$ defined as:

$$E(K) = \lambda \sum_q E_{B,\Phi}(K_q) + \sum_{(p,q)} E_\Phi(K_q, K_p) \quad (10)$$

where: $E_{B,\Phi}(K_q) = \log p(B_q|D_q)$ (from (7)); $E_\Phi(K_q, K_q) = \log P(K_q - K_p|\Phi_q, \Phi_p)$ (from (9)); and (p, q) represent pairs of neighboring pixels. Inference is performed by loopy belief propagation as described in [**26**] [**27**] using a simultaneous message passing schedule, with an 8-connected neighborhood. We cease iterations when one of the convergence criteria are met: either the change in total sum log marginalized probability is below a threshold, or there is no change in labels for a set number of iterations.

4. EXPERIMENTS

Our algorithm was tested on a set of 45 images (consisting of wrapped phase Φ and brightness B) collected from 3 different locations: a home, an office setting, and a computer lab. Only indoor scenes—the primary setting for ToF sensors—were chosen to be included as a known issue for active illumination sensors is excessive ambient light. We attempted to capture scenes with a variety depths, so to include a range of difficulties for which to test the algorithm. Data was captured using a prototype ToF camera with a resolution of 320x200 pixels, on loan from a commercial vendor.

Each static scene was captured at two frequencies, 51.4MHz and 68.6MHz, which determined unambiguous ranges of 2.2m and 2.9m respectively. Ground truth was determined by combining these pairs of captures using the approach of [**2**], which uses both frequencies to obtain an unambiguous phase measurement. We ran our phase unwrapping algorithm on both frequencies individually. In addition, we tested our algorithm on an extended range of modulation frequencies by synthesizing wrapped phase data from the data at the two original modulation frequencies. This was obtained starting from the "ground truth" depth data (obtained as described above), and manually wrapping the phase according to the desired modulation frequency. We created synthetic phase images with frequencies ranging from 50Mhz to 150Mhz (unambiguous range of 1 - 3m).

We compared the results of our algorithm against the method of Choi et al. [**5**], which is considered the state of the art for single frequency solutions. This is a natural comparison, as both methods incorporate the intensity image. The key difference is that in [**5**] each pixel is classified as bright or dim based on the distribution of intensity in that image. Thus the contribution from the intensity image is binary in nature, supporting each pixel as near (0 wraps) or far (1 or more wraps). Since the original implementation of Choi's algorithm was not available to us (due to proprietary restrictions), we implemented it in Matlab following the algorithm's description in [**5**].

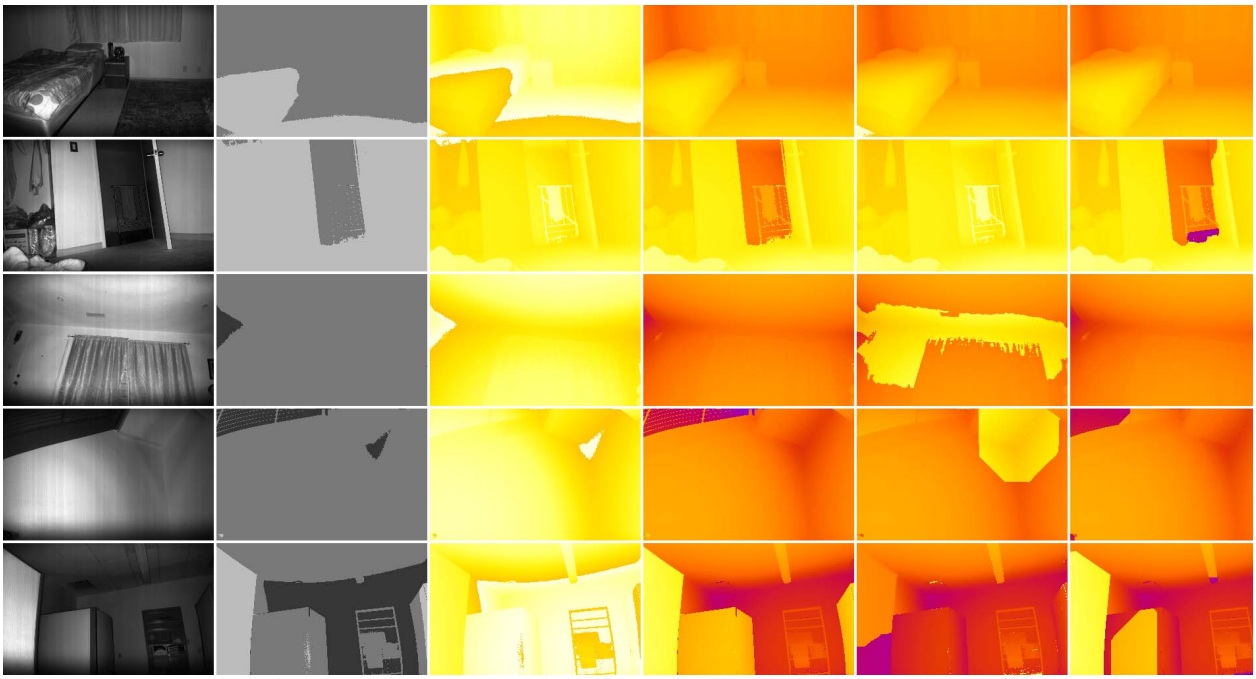

Figure 3. Selected scenes from Experiment 1. The top row is taken from the *easy* group, second and third rows are from the medium and the bottom two from *hard*. The first column shows the intensity image from the active illumination. The second column is the ground truth number of phase wraps, from 0 to 3 wraps with darker greys being more wraps. The 3rd through 5th columns show phase, increasing from hot white to cool purple. The 3rd column shows the measured phase (wrapped at 2π). The fourth column is representation of the unwrapped phase or depth, as measured by a multi-frequency ToF camera. The final two columns compare the best solutions from the method by [5] to our own algorithm at the end.

We divided our data by the maximum number of phase wraps observed in each scene, either 1, 2, or 3, (that is up to 4× the unambiguous range). The data is further subdivided by frequency in Figure 4, where we summarize the performance of the algorithms. Alongside we present the performance of our smoothness and brightness terms alone. For the smoothness term alone, we set $\lambda = 0$ and run the loopy belief propagation. For the brightness term, rather than using belief propagation, we simply choose the value of K that maximizes $p(B|K, \Phi)$, as defined in (7).

In the 14 *easy* test cases (up to 1 phase wrap), our method outperformed Choi's method with average of 99.4% correctly labeled pixels compared to 98.3%. A selection of results is displayed in Figure 3 (first three rows). For example, the first row of Figure 3 shows a simple bedroom scene. Note that our algorithm is able to recover the section of floor in the bottom left corner beneath the bed. In the second row of Figure 3,our algorithm is able to correctly identify the small area through the door as being farther away.

For the 45 scenes with 2 phase wraps, both algorithms have a tougher time, but our advantages over Choi start to become more apparent: successfully labelling still 91.4% of the pixels versus 87.1%. Examples of these scenes are shown in Figure 3, last two rows. Careful analysis shows that the key cause for this bad performance was due to a misclassification of lighter and darker pixels as nearer and farther in the early stage of Choi's algorithm. For example, in the case shown in the last row of Figure 3, the low albedo of the television set results in dark pixels that are incorrectly classified as far. Similar pitfalls are avoided with our algorithm, as dark pixels are considered uninformative under our image brightness model (see Figure 2).

In the hardest of cases, with 3 wraps, the gap continues to widen, with 83.3% against Choi's 73.8%. This trend goes on as is demonstrated by increasing the maximum number of wraps by synthesizing cases using the same depth and brightness values, but computing new values for Φ as they would appear for higher frequencies.

The results of experiments with synthetically produced phase images over varying modulation frequency are shown in Figure 5. We tested at 7 different simulated frequencies for all 45 test scenes at half-resolution. As the frequency increases (and the unambiguous range decreases), the phase unwrapping problem becomes more challenging, given the larger number of possible wrap values for the same maximum distance. Both methods show a decrease in performance for increasing modulation frequency, although this decrease is relatively minor with our method. Although the method is presented as applying to multiple wraps,

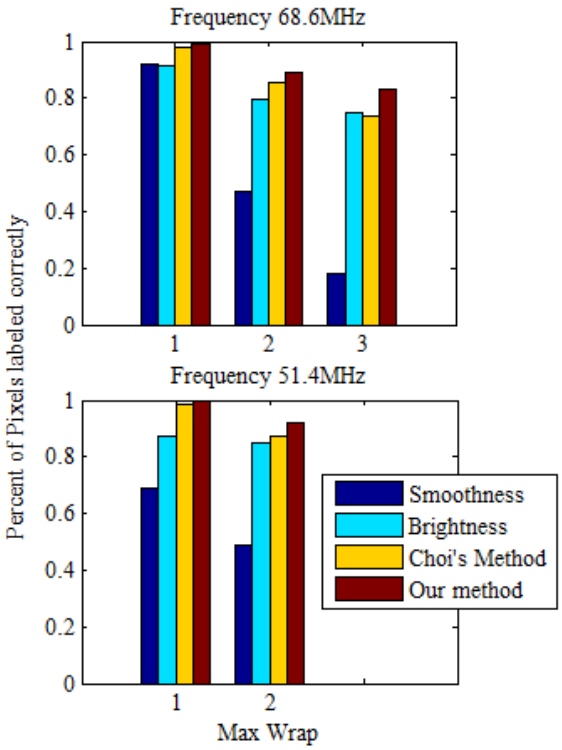

Figure 4. Results on real data. These bar graphs compare, for each frequency, the performance of the individual terms of our proposed solution, Smoothness (blue) and Brightness (cyan), alongside the method proposed by Choi et al. (yellow) and our complete proposed solution (red). The data is divided by the maximum number of wraps, as the phase unwrapping problem becomes more difficult with a larger range of phase. Note that for 51.4MHz, there were never more than 2 phase wraps.

Choi's algorithm clearly fails in the case of high modulation frequency.

We have implemented the loopy belief propagation algorithm in Matlab. For the 90 scenes worth of real data, a grid search was performed over the parameters λ and σ to find the best performance in terms of correctly labeled pixels. The algorithm's convergent criterea to be a change in negative log marginal probability of less than 1e-10 or if no change in label after 4 consecutive iterations. We found each scene to converge after an average of 330 iterations, with an average time of about .5 sec per iteration.

5. CONCLUSIONS

This paper presented an approach for phase unwrapping of time-of-flight systems with the potential to improve frame rates and avoid motion artifacts. The wrap number is estimated directly, which guarantees no residues and provides an exact solution, as opposed to a solution "up to a constant." Our method makes use of valuable information from the signal intensity, without

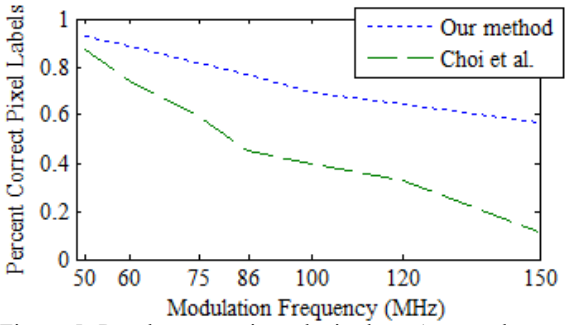

Figure 5. Results on semi-synthetic data. Averaged over all 45 scenes at each frequency, regardless of the max wrap value. The maximum wrap value reached (at 150MHz) was 8.

resorting to early-commitment, segmentation-based procedure as with other previous approaches. The experimental results show that the proposed algorithm outperforms existing methods that also use brightness information. Future work will look into adding additional priors such as local smoothness of albedo and surface normal, which may easily be integrated in our MRF framework.

REFERENCES

[1] Thierry Oggier et al., "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger)," in *Optical Design and Engineering*, 2004, p. 534.

[2] S Burak Göktürk, Hakan Yalcin, and Cyrus Bamji, "A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions," in *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 3*, Washington, DC, USA, 2004, p. 35.

[3] G. J. Iddan and G. Yahav, "3D Imaging in the studio (and elsewhere)," in *Proc. SPIE 4298, Three-Dimensional Image Capture and Applications*, 2001.

[4] Dean Takahashi, "Microsoft discloses details on Xbox One and Kinect chips," *VentureBeat*, August 2013.

[5] Ouk Choi et al., "Range unfolding for time-of-flight depth cameras," in *International Conference on Image Processing (ICIP)*, Hong Kong, 2010.

[6] Dennis Ghiglia and Mark Pritt, *Two-dimensional phase unwrapping: theory, algorithms, and software*. New York: Wiley, 1998.

[7] Brendan J. Frey, Ralf Koetter, and Nemanja Petrovic, "Very loopy belief propagation for unwrapping phase images," in *Neural Information Processing Systems (NIPS)*, Vancouver, 2001.

[8] D. Droeschel, D. Holz, and S. Behnke, "Probabilistic phase unwrapping for time-of-flight cameras," in

41st International Symposium on Robotics (ISR), Munich, Germany, 2010, pp. 1-7.

[9] W. Xu et al., "Phase-unwrapping of SAR interferogram with multi-frequency or multi-baseline," in *Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 1994.

[10] G., and Bioucas-Dias, J. Valadao, "Phase imaging: Unwrapping and denoising with diversity and multi-resolution," in *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, Lausanne, Switzerland, 2008.

[11] D. Droeschel, D. Holz, and S. Behnke, "Multi-frequency phase unwrapping for time-of-flight cameras," in *Intelligent Robots and Systems (IROS)*, Taipei, 2010.

[12] D. Falie and V. Buzuloiu, "Wide range time of flight camera for outdoor surveillance," in *Microwaves, Radar and Remote Sensing Symposium (MRRS)*, Kiev, 2008.

[13] O. Choi, S. Lee, and H Lim, "Interframe consistent multifrequency phase unwrapping for time-of-flight cameras," *Optical Engineering*, vol. 52, no. 5, 2013.

[14] J. Mei, A. Kirmani, A. Colaco, and V.K. Goyal, "Phase unwrapping and denoising for time-of-flight imaging using generalized approximate message passing," in *20th IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, 2013, pp. 364-368.

[15] A. D. Payne, A. P. Jongenelen, A. A. Dorrington, M. J. Cree, and D. A Carnegie, "Multiple frequency range imaging to remove measurement ambiguity," in *Conference on Optical 3-D Measurement Techniques*, Vienna, Austria, 2009.

[16] Sigurjon Arni Gudmundsson, Henrik Aanaes, and Rasmus Larsen, "Fusion of stereo vision and time-of-flight imaging for improved 3d estimation," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3, pp. 425-433, 2008.

[17] C. Beder, B. Barzak, and R. Koch, "A combined approach for estimating patchlets afrom pmd depth images and stereo intensity images," in *German Association for Pattern Recognition (DAGM)*, Heidelberg, 2007.

[18] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," in *Computer Vision and Pattern Recognition*, Anchorage, 2008.

[19] C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, "A probabilistic approach to ToF and stereo data fusion," in *3DPVT*, Paris, 2010.

[20] Ouk Choi and S. Lee, "Fusion of time-of-flight and stereo for disambiguation of depth measurements," in *Asian Conference on Computer Vision*, Daejeon, Korea, 2012, pp. 640-653.

[21] Ouk Choi and Seungkyu Lee, "Wide range stereo time-of-flight camera," in *International Conference on Image Processing*, Orlando, 2012.

[22] Martin Böhme, Martin Haker, Thomas Martinetz, and Erhardt Barth, "Shading constraint improves accuracy of time-of-flight measurements," *Computer vision and image understanding*, vol. 114, no. 12, pp. 329-1335, 2010.

[23] Shane H. McClure, Cree Michael J., Adrian Dorrington, and Andrew Payne, "Resolving depth-measurement ambiguity with commercially available range imaging cameras," in *IS&T/SPIE Electronic Imaging*, San Jose, CA, USA, 2010.

[24] Serban Oprisescu, Dragos Falie, Mihai Ciuc, and Vasile Buzuloiu, "Measurements with ToE cameras and their necessary corrections," in *International Symposium on Signals, Circuits and Systems*, Iasi, Romania, 2007.

[25] A. Kadambi et al., "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles," *ACM Transactions on Graphics*, vol. 32, no. 6, p. 167, 2013.

[26] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient belief propagation for early vision," *International journal of computer vision*, vol. 70, no. 1, pp. 41-54, 2006.

[27] R. Szeliski et al., "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 1068-1080, 2008.